

CLAIMS:

What is claimed is:

- DC*
1. A method of workload management in a data processing system, comprising:
5 classifying a process into a class of a plurality of predefined classes, the plurality of predefined classes having corresponding system resource shares and tiers;
10 performing workload management with respect to classes in other tiers based on priorities of the tiers; and
15 performing workload management with respect to other classes within a same tier as the class into which the process is classified, based on the corresponding system resource shares of the class and of the other classes.
2. The method of claim 1, wherein the system resource shares are used to determine a percentage of system resources to be allocated to the process in the class relative to other processes in other classes.
20
3. The method of claim 1, wherein the system resource tiers provide a measure of importance of a set of classes relative to another set of classes.
25
4. The method of claim 1, wherein the process is classified into a class based on one or more classification rules.
30
5. The method of claim 1, further comprising storing the process in a workload queue based on the classification of the process into the class.

PCT/US2000/026800

6. The method of claim 1, wherein each class of the plurality of predefined classes has a minimum and a maximum resource limit, the minimum and maximum resource limits defining a minimum and a maximum amount of a resource that may be allocated to the class as a percentage of the resource.
7. The method of claim 1, wherein the system resource tiers are organized hierarchically by priority.
8. The method of claim 1, wherein performing workload management with respect to other classes within a same tier comprises determining a percentage goal for the process as a function of a number of system resource shares associated with the class in which the process is classified divided by a total number of shares allocated to active classes in the same tier as the class in which the process is classified.
9. The method of claim 1, wherein each process in each of the plurality of predefined classes has an associated priority component that is used, along with the system resource shares of the class, to calculate a resource allocation priority.
10. The method of claim 1, wherein performing workload management with respect to other classes within the same tier comprises determining a resource allocation priority range for the class based on a minimum resource usage limit, a maximum resource usage limit, and a resource usage goal.

DO NOT REMOVE

11. The method of claim 10, wherein performing workload management with respect to other classes within the same tier further comprises adjusting the resource allocation priority range for the class based on a comparison of an
5 actual amount of the resource being utilized by the class to the maximum resource usage limit, the minimum resource usage limit, and the resource usage goal.
12. The method of claim 8, wherein if the percentage
10 goal is below a minimum resource usage limit, the class is favored for additional usage of the system resource, and wherein if the calculated percentage goal is above a maximum resource usage limit, the class is not favored for additional usage of the system resource.
15
13. The method of claim 12, wherein the minimum resource usage limit and maximum resource usage limit are retrieved from a share/tier profile storage.
20
14. The method of claim 1, wherein each of the plurality of predefined classes includes an associated absolute maximum resource usage limit, and wherein if the resource allocation for processes in a class exceed an absolute maximum resource usage limit for the class, additional
25 resource allocation to the class is suspended.
15. The method of claim 1, further comprising determining a class priority adjustment for each of the plurality of predefined classes, wherein the class
30 priority adjustment is determined based on a delta value computed for every predetermined time increment.

00000000000000000000000000000000

16. The method of claim 15, wherein the delta value is a difference between an average resource usage over a specified time interval and a resource usage for a last predetermined time increment.

5

17. The method of claim 10, further comprising adjusting a priority of the class by a class priority adjustment value, the class priority adjustment value being determined based on a comparison of actual resource usage 10 to the minimum resource usage limit, the maximum resource usage limit, and a resource usage goal.

18. The method of claim 17, wherein if the actual resource usage is between the maximum resource usage 15 limit and an absolute maximum resource usage limit, the class priority adjustment is set to disfavor the class from being allocated additional amounts of the resource.

19. The method of claim 17, wherein if the actual 20 resource usage is between the resource usage goal and the maximum resource usage limit and a difference between an average resource usage over a specified time interval and a resource usage for a last predetermined time increment is less than zero, the class priority 25 adjustment is not changed.

20. The method of claim 17, wherein if the actual resource usage is between the minimum resource usage 30 limit and the resource usage goal and a difference between an average resource usage over a specified time interval and a resource usage for a last predetermined time increment is greater than zero, the class priority

DOCKET NO. AUS9-2000-0268-US1

adjustment is incremented by a function of the difference.

21. The method of claim 17, wherein if the actual
5 resource usage is below the resource usage goal and the
difference between an average resource usage over a
specified time interval and a resource usage for a last
predetermined time increment is greater than zero, the
class priority adjustment is not changed.

10

22. The method of claim 17, wherein if the actual
resource usage is below the resource usage goal and the
difference between an average resource usage over a
specified time interval and a resource usage for a last
15 predetermined time increment is less than or equal to
zero, the class priority adjustment is decremented by a
function of the difference.

23. The method of claim 7, wherein processes in classes
20 in lower priority tiers are only able to obtain access to
a resource if processes in classes in higher priority
tiers do not use all of the resource or if the processes
in each of the classes in the higher priority tier have
reached an absolute maximum resource usage limit.

25

24. A computer program product in a computer readable
medium for workload management in a data processing
system, comprising:
first instructions for classifying a process into a

30 class of a plurality of predefined classes, the plurality
of predefined classes having corresponding system
resource shares and tiers;

second instructions for performing workload

DOCKET NO. AUS9-2000-0268-US1

management with respect to classes in other tiers based on priorities of the tiers; and

- third instructions for performing workload management with respect to other classes within a same tier as the class into which the process is classified, based on the corresponding system resource shares of the class and of the other classes.

25. The computer program product of claim 24, wherein
10 the system resource shares are used to determine a percentage of system resources to be allocated to the process in the class relative to other processes in other classes.

15 26. The computer program product of claim 24, wherein the system resource tiers provide a measure of importance of a set of classes relative to another set of classes.

27. The computer program product of claim 24, wherein
20 the process is classified into a class based on one or more classification rules.

28. The computer program product of claim 24, further comprising fourth instructions for storing the process in
25 a workload queue based on the classification of the process into the class.

29. The computer program product of claim 24, wherein each class of the plurality of predefined classes has a
30 minimum and a maximum resource limit, the minimum and maximum resource limits defining a minimum and a maximum amount of a resource that may be allocated to the class as a percentage of the resource.

000456789000

30. The computer program product of claim 24, wherein the system resource tiers are organized hierarchically by priority.

5

31. The computer program product of claim 24, wherein the third instructions for performing workload management with respect to other classes within a same tier comprises instructions for determining a percentage goal for the process as a function of a number of system resource shares associated with the class in which the process is classified divided by a total number of shares allocated to active classes in the same tier as the class in which the process is classified.

10

32. The computer program product of claim 24, wherein each process in each of the plurality of predefined classes has an associated priority component that is used, along with the system resource shares of the class, 20 to calculate a resource allocation priority.

15

33. The computer program product of claim 24, wherein the third instructions for performing workload management with respect to other classes within the same tier 25 comprises instructions for determining a resource allocation priority range for the class based on a minimum resource usage limit, a maximum resource usage limit, and a resource usage goal.

25

34. The computer program product of claim 33, wherein the third instructions for performing workload management with respect to other classes within the same tier further comprises instructions for adjusting the resource

P A G E - 0 9

allocation priority range for the class based on a comparison of an actual amount of the resource being utilized by the class to the maximum resource usage limit, the minimum resource usage limit, and the resource usage goal.

35. The computer program product of claim 31, wherein if the percentage goal is below a minimum resource usage limit, the class is favored for additional usage of the system resource, and wherein if the calculated percentage goal is above a maximum resource usage limit, the class is not favored for additional usage of the system resource.

15 36. The computer program product of claim 35, wherein the minimum resource usage limit and maximum resource usage limit are retrieved from a share/tier profile storage.

20 37. The computer program product of claim 24, wherein each of the plurality of predefined classes includes an associated absolute maximum resource usage limit, and wherein if the resource allocation for processes in a class exceed an absolute maximum resource usage limit for 25 the class, additional resource allocation to the class is suspended.

30 38. The computer program product of claim 24, further comprising fourth instructions for determining a class priority adjustment for each of the plurality of predefined classes, wherein the class priority adjustment is determined based on a delta value computed for every

DO NOT FILE - CONFIDENTIAL

predetermined time increment.

39. The computer program product of claim 38, wherein
the delta value is a difference between an average
5 resource usage over a specified time interval and a
resource usage for a last predetermined time increment.

40. The computer program product of claim 33, further
comprising fourth instructions for adjusting a priority
10 of the class by a class priority adjustment value, the
class priority adjustment value being determined based on
a comparison of actual resource usage to the minimum
resource usage limit, the maximum resource usage limit,
and a resource usage goal.
15

41. The computer program product of claim 40, wherein if
the actual resource usage is between the maximum resource
usage limit and an absolute maximum resource usage limit,
the class priority adjustment is set to disfavor the
20 class from being allocated additional amounts of the
resource.

42. The computer program product of claim 40, wherein if
the actual resource usage is between the resource usage
25 goal and the maximum resource usage limit and a
difference between an average resource usage over a
specified time interval and a resource usage for a last
predetermined time increment is less than zero, the class
priority adjustment is not changed.
30

43. The computer program product of claim 40, wherein if
the actual resource usage is between the minimum resource

0001447525425475000

usage limit and the resource usage goal and a difference between an average resource usage over a specified time interval and a resource usage for a last predetermined time increment is greater than zero, the class priority adjustment is incremented by a function of the difference.

44. The computer program product of claim 40, wherein if the actual resource usage is below the resource usage goal and the difference between an average resource usage over a specified time interval and a resource usage for a last predetermined time increment is greater than zero, the class priority adjustment is not changed.

15 45. The computer program product of claim 40, wherein if the actual resource usage is below the resource usage goal and the difference between an average resource usage over a specified time interval and a resource usage for a last predetermined time increment is less than or equal 20 to zero, the class priority adjustment is decremented by a function of the difference.

46. The computer program product of claim 30, wherein processes in classes in lower priority tiers are only able to obtain access to a resource if processes in 25 classes in higher priority tiers do not use all of the resource or if the processes in each of the classes in the higher priority tier have reached an absolute maximum resource usage limit.

30

47. A workload management apparatus in a data processing system, comprising:

a process classifier for classifying a process into

DOCUMENT EDITION

a class or a plurality of predefined classes, the plurality of predefined classes having corresponding system resource shares and tiers; and

- 5 a processor coupled to the process classifier for performing workload management with respect to classes in other tiers based on priorities of the tiers, and performing workload management with respect to other classes within a same tier as the class into which the process is classified, based on the corresponding system
10 resource shares of the class and of the other classes.

48. The workload management apparatus of claim 47, wherein the system resource shares are used to determine a percentage of system resources to be allocated to the
15 process in the class relative to other processes in other classes.

49. The workload management apparatus of claim 47, wherein the system resource tiers provide a measure of
20 importance of a set of classes relative to another set of classes.

50. The workload management apparatus of claim 47, further comprising a classification rules storage device
25 for storing classification rules, wherein the process classifier classifies the process into a class based on one or more classification rules.

DOCKET NUMBER